TESTING PROBLEMS WITH ADDITIONAL OR MISSING DATA

By

Morris Eaton[1] and Takeaki Kariya

University of Minnesota

Technical Report No. 370

February 29, 1980

1. This research was supported in part by a National Science Foundation Grant MCS 7725112.

ABSTRACT

Consider data on p coordinates which are partitioned into groups of p_1 and p_2 coordinates. Consider n p-dimensional observation vectors and m_i p_i-dimensional observation vectors, $i = 1,2$ where all $n+m_1+m_2$ vectors are independent. Assume a $N_p(\mu,\Sigma)$ distribution for the p-coordinates and write $\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$ and $\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$ where μ_i is $p_i \times 1$, $i = 1,2$ with Σ partitioned similarly. In this paper we consider testing $H_0^{(1)}$: $\mu_1 = 0$ and $H_0^{(2)}$: $\mu = 0$ when $m_2 = 0$. For $H_0^{(1)}$, the LRT is UMP invariant and for $H_0^{(2)}$, no locally most powerful invariant (LMPI) test exists. For $H_0^{(3)}$: $\Sigma_{12} = 0$, we derive a LMPI test and show that when $m_1 = 0$ and $m_2 > 0$, this test differs from the LRT.

American Mathematical Society 1970 Subject
Classifications: Primary 62H15
Secondary 62A05

Key Words and Phrases: missing data, locally most powerful invariant tests, testing independence, invariance.

§0: Introduction

Because of their common occurence in practice, there has been a continuing interest in inference problems where there is missing or extra data. The causes for the data to be missing or extra will not be discussed explicitly in this paper, but will be implicit in our assumptions concerning the likelihood function of the data (see Section 1). For an illuminating discussion of such issues, the reader is referred to Rubin (1963). With the likelihood assumed in (1.1), it is equivalent to think of certain parts of the data as additional or the "complementary" parts of these data as missing.

The problems treated in this paper concern data on p coordinates which are partitioned into two groups of p_1 and p_2 coordinates - so $p_1+p_2 = p$ and $1 \leq p_i < p$, $i = 1,2$. It is assumed that we have n p-dimensional observation vectors and m_i p_i-dimensional observation vectors, $i = 1,2$. All $n+m_1+m_2$ vectors are assumed to be independent. Thus, there are n "complete" observations, m_1 "extra" observations on the first p_1 coordinates, and m_2 "extra" observations on the last p_2 coordinates. When m_1 (or equivalently m_2) is zero, then the data is in triangularly partitioned form. Under the assumption of multivariate normality, Bhargava (1962) derived maximum likelihood estimators (MLE's) and likelihood ratio tests (LRT's) for a number of problems when the data has a general triangular form. This triangular form permits the explicit calculation of MLE's and LRT's along with the relevant distribution theory. Morrison and Bhoj (1973) discuss the power of the LRT for testing a mean vector is zero when $m_1 = 0$.

Ordinarily, likelihood methods are proposed for problems with missing data - especially when the normal distribution is involved. However, in some situations, the likelihood equations cannot be solved explicitly. The article by Hartley and Hocking (1971) provides a good overview of the subject and an extensive bibliography. The recent work of Little (1976) is concerned solely with the normal distribution but general patterns of missing data are allowed. Little compares a variety of estimators both asymptotically and numerically.

To illustrate the possible difficulties involved in missing data problems, consider the following case: $p_1 = p_2 = 1$, the "complete" data is a sample of n from a bivariate normal distribution with unknown mean vector $\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$ and unknown covariance matrix, and the two extra samples are from univariate normal populations with means and variances of the marginal distributions of the bivariate normal. Suppose the problem is to test that $\mu_1 = \mu_2$. If $n = 0$, this problem is the Behrens-Fisher problem. When $n > 0$, the problem should be no easier than the Behrens-Fisher problem and the work thus far justifies this belief. A comparison of several different proposals to solve this problem is given in Ekbohm (1976).

The purpose of the present paper is to discuss the existence or nonexistence of tests with certain optimum properties. In Section 1, we set notation and derive a canonical form for the data under consideration. It is assumed that n p-dimensional normal (μ,Σ) vectors are available and m_i p_i-dimensional normal (μ_i,Σ_{ii}), $i = 1,2$, are available where

$$\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \quad \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$$

with μ_i: $p_i \times 1$ and Σ_{ij}: $p_i \times p_j$, $i,j = 1,2$. All the parameters are assumed unknown.

In Section 2, it is assumed that $m_2 = 0$ so no "extra" data is available on the last p_2-coordinates. For the problem of testing $\mu_1 = 0$ versus $\mu_1 \neq 0$, the LRT is shown to be uniformly most powerful invariant. However, for testing $\mu = 0$ versus $\mu \neq 0$, a locally most powerful invariant test does not exist. A few comments concerning the LRT of $\mu = 0$ are given.

In Section 3, we consider the problem of testing $\Sigma_{12} = 0$ when both m_1 and m_2 are non-negative. In this case we derive a locally most powerful invariant test. When $m_1 = 0$ and $m_2 > 0$ (or $m_1 > 0$ and $m_2 = 0$), this test is different from the LRT. (When $m_1 > 0$ and $m_2 > 0$, the LRT is not known explicitly). In fact, the LRT does not utilize the "extra" data at all and is identical to the LRT when $m_1 = m_2 = 0$. This point is discussed and we propose a possible test statistic for testing $\Sigma_{12} = 0$ which utilizes the additional information. Two examples are presented in Section 4.

The missing data patterns considered in this paper are among the most simple, but our results indicate the variety of possible answers one can obtain when: (i) comparing LRT's to optimum (in some sense) tests when they both exist and (ii) trying to settle questions concerning the existence of optimum tests. Invariance considerations play a central role in this paper. Rather than using sufficiency and invariance to reduce the available data, we employ the method of averaging over groups to obtain the density function of a maximal invariant. Of course, we have only been able to employ this technique when the group under consideration acts transitively on the null hypothesis. The particular representation result we have used is due to Andersson (1979) and this result is outlined in Appendix I. We have found Andersson's approach easier to apply than the approach of Wijsman (1967). The application of Andersson's result to the proof of Theorem 3.1 is given in Appendix II. Some of the details are only sketched as they are similar to those in Schwartz (1967). The proofs of Theorems 2.1 and 2.2 are omitted as they are similar to that outlined for Theorem 3.1.

§1: Notation and A Canonical Form

The extra (or missing) data problem to be considered here is one of the simplest but illustrates the mathematical problems encountered when dealing with such models. Consider a multivariate normal population of dimension p with a mean vector μ (a column vector) and a $p \times p$ non-singular covariance matrix Σ. Write $p = p_1 + p_2$ where $1 \leq p_i < p$ for $i = 1,2$ and partition μ and Σ as

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

with μ_i being $p_i \times 1$ and Σ_{ij} being $p_i \times p_j$ for $i,j = 1,2$. It is assumed that we have "complete" observations $X_1,\cdots,X_n$ which are i.i.d. $N_p(\mu,\Sigma)$ and "marginal" observations $X_{i1},\cdots, X_{im_i}$, $i = 1,2$ which are i.i.d. $N_{p_i}(\mu_i,\Sigma_{ii})$ for $i = 1,2$. In terms of data matrices, the complete sample yields $\tilde{X}$: $n \times p$ with rows X_i', $i = 1,\cdots,n$ so

$$\mathcal{L}(\tilde{X}) = N(e_n\mu', I_n \otimes \Sigma)$$

where e_n is the vector of ones in R^n. Here, the notation "$\mathcal{L}(\cdot)$" means the distribution of "$\cdot$" and $\otimes$ denotes the Kronecker product.

Similarly, the marginal samples yield data matrices $\tilde{X}_1:m_1\times p_1$ and $\tilde{X}_2:m_2\times p_2$ with

$$L(\tilde{X}_i) = N(e_{m_i}\mu_i', I_{m_i}\otimes\Sigma_{ii}),\ i = 1,2.$$

It is convenient to transform the data $\tilde{X},\tilde{X}_1$ and $\tilde{X}_2$ into what will be called the canonical form. Let Γ be an $n\times n$ orthogonal matrix with first row $e_n'/\sqrt{n}$. Then the transpose of the first row of the matrix $\Gamma\tilde{X}$ has a $N(\sqrt{n}\mu,\Sigma)$ distribution and is independent of the remaining $(n-1)$ rows which are i.i.d. $N(0,\Sigma)$. Let $U\in R^p$ be the transpose of the first row of $\Gamma\tilde{X}$ multiplied by $1/\sqrt{n}$ and let $V:(n-1)\times p$ be the remaining $(n-1)$ rows of $\Gamma\tilde{X}$. Then U and V are independent with

$$L(U) = N(\mu, \tfrac{1}{n}\Sigma)$$

and

$$L(V) = N(0, I_{(n-1)}\otimes\Sigma).$$

Transforming $\tilde{X}_1$ in a similar manner leads to $U_1\in R^{p_1}$ and $V_1:(m_1-1)\times p_1$ which are independent and satisfy

$$L(U_1) = N(\mu_1, \tfrac{1}{m_1}\Sigma_{11}),\ L(V_1) = N(0, I_{(m_1-1)}\otimes\Sigma_{11}).$$

Similarly, transforming $\tilde{X}_2$ leads to $U_2\in R^{p2}$ and $V_2:(m_2-1)\times p_2$ which are independent and satisfy

$$L(U_2) = N(\mu_2, \tfrac{1}{m_2}\Sigma_{22}),\ L(V_2) = N(0, I_{(m_2-1)}\otimes\Sigma 22).$$

In summary, the complete and partial data $\tilde{X},\tilde{X}_1$ and $\tilde{X}_2$ can be relabeled to yield U,U_1,U_2 and V,V_1,V_2 with the given distributions.

With the above discussion in mind, we now describe the canonical form for our extra data problem. The mean vector μ and covariance matrix Σ, partitioned as

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

and the dimension parameter $p = p_1+p_2$ are as before. Consider independent random vectors $Y \in R^p$, $X_i \in R^{p_i}$, $i = 1, 2$ and independent random matrices $V{:}n\times p$, $V_i{:}m_i\times p_i$, $i = 1, 2$ such that

$$(1.1) \quad \begin{cases} L(Y) = N(\mu, c\Sigma) \\ L(X_i) = N(\mu_i, c_i\Sigma_{ii}), \ i = 1,2 \\ L(V) = N(0, I_n \otimes \Sigma) \\ L(V_i) = N(0, I_{m_i} \otimes \Sigma_{ii}), \ i = 1,2 \end{cases}$$

where c, c_1 and c_2 are known positive constants, and n, m_1 and m_2 are known positive integers. Observations which are represented in the form (1.1) will be said to be in canonical form. For some of the problems treated below, the data is assumed to be in canonical form. However, to properly motivate the examples in Section 4, it is necessary to describe the original problem before transforming it to the form (1.1).

In some cases, parts of the data in (1.1) will be missing. For example, if there is no marginal sample on the last p_2-coordinate, then both X_2 and V_2 are missing in (1.1) and $m_2 = 0$. In fact, this will be the case considered in the next section where we take up the problem of testing hypotheses about μ. The full generality of (1.1) is used in Section 3 for testing that $\Sigma_{12} = 0$.

Throughout this paper, it is assumed that sample sizes are large enough so maximum likelihood estimators exist - that is, we assume $\min\{n+m_1, n+m_2\} > \max\{p_1, p_2\}$ in the canonical model (1.1). Invariance will be a central theme in this paper and $G\ell_p$ will always denote the group of $p\times p$ non-singular real matrices.

§2: Tests on Means

In this section, we consider data Y, X_1, V and V_1 in the canonical form (1.1) where X_2 and V_2 are not present. As remarked earlier, this means that no "extra" data was available on the last p_2 coordinates of our basic sample. Of course, it is assumed that Y, X_1, V and V_1 are mutually independent and

$$(2.1) \quad \begin{cases} \mathcal{L}(Y) = N(\mu, c\Sigma) \\ \mathcal{L}(X_1) = N(\mu_1, c_1\Sigma_{11}) \\ \mathcal{L}(V) = N(0, I_n \otimes \Sigma) \\ \mathcal{L}(V_1) = N(0, I_{m_1} \otimes \Sigma_{11}) \end{cases}$$

in the notation of Section 1. Based on the data (2.1), we now want to discuss the problem of testing $H:\mu_1 = 0$ versus $H_1:\mu_1 \neq 0$. This testing problem is invariant under a group of transformations acting on the sample space. In particular, consider the group G whose elements are $g = (A,a)$ with $A \in G_1$ and $a \in R^p$ where

$$A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \; A_{ii} \in G\ell_{p_i}, \; i=1,2$$

and

$$a = \begin{pmatrix} 0 \\ a_2 \end{pmatrix}; \; a_2 \in R^{p_2}.$$

The action of $g = (A,a)$ on a sample point (Y, X_1, V, V_1) is

$$g(Y, X_1, V, V_1) = (AY + a, A_{11}X_1, VA', V_1A'_{11})$$

and the corresponding action on the parameter point (μ,Σ) is

$$g(\mu,\Sigma) = (A\mu+a,\ A\Sigma A') .$$

The composition of two group elements is

$$(A,a)(B,b) = (AB,Ab+a).$$

It is now a routine matter to check that the testing problem is invariant under the group G. A maximal invariant in the parameter space is

$$\delta_1 = \mu_1'\Sigma_{11}^{-1}\mu_1 .$$

In terms of δ_1, the null hypothesis is $H_0:\delta_1 = 0$ and the alternative is $H_1:\delta_1 > 0$. The next result will allow us to derive a uniformly most powerful invariant (UMPI) test of H_0 versus H_1.

<u>Theorem 2.1</u>: Let P_{δ_1} denote the probability measure of a maximal invariant at the parameter value δ_1. Then the Radon-Nibodym deivative dP_{δ_1}/dP_0 is given by

$$(2.2)\qquad R(t|\delta_1) = \sum_{j=0}^{\infty} \mathcal{E}_j\delta_1^j t^j$$

where the constants $\mathcal{E}_j$ are positive and

$$(2.3)\qquad t \equiv k\left(\frac{Y_1}{c} + \frac{X_1}{c_1}\right)' \left(S_{11}+V_1' V_1 + \frac{Y_1Y_1'}{c} + \frac{X_1X_1'}{c_1}\right)^{-1} \left(\frac{Y_1}{c} + \frac{X_1}{c_1}\right).$$

Here,

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},\ Y_i \in R^{p_i},\ i = 1,2\ ,\quad k^{-1} = \frac{1}{c} + \frac{1}{c_1}\ ,$$

and

$$V'V \equiv S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

with S_{ij} of dimension $p_i \times p_j$, $i,j = 1,2$.

Proof: The proof of this result is similar to the proof of Theorem 3.1 which is outlined in Appendix II.

Corollary 1.1: The test of $H_0:\delta_1 = 0$ versus $H_1:\delta_1 > 0$ which rejects for large values of t given by (2.3) is UMPI.

Proof: The density function (with respect to P_0) of a maximal invariant is $R(\cdot|\delta_1)$ when the parameter value is δ_1, and $R(\cdot|0) \equiv 1$. Since $R(\cdot|\delta_1)$ is an increasing function of t, the most powerful level α test based on t of $H_0:\delta_1 = 0$ versus $\tilde{H}_1:\delta_1 = \delta^*$ rejects for large values of t. Since this test does not depend on δ^*, this implies the desired result.

It is not difficult to show that the likelihood ratio test (LRT) of $H_0:\mu_1 = 0$ versus $H_1:\mu_1 \neq 0$ is equivalent to the test which rejects for large values of t. Standard arguments show that the statistic t has a central Beta distribution under H_0 and a non-central Beta distribution under H_1. See Bhargava (1962) for details.

We now turn to the problem of testing $H_0:\mu = 0$ versus the alternative $H_1:\mu \neq 0$. In this case, the situation is substantially different from the first case considered. Again, we take the data of the problem to be given by (2.1). This testing problem is invariant under the group G_0, a subgroup of G, defined by

$$G_0 = \{g = (A,a) | g \in G,\ a = 0\}.$$

Thus, an element of G_0 is simply a $p \times p$ matrix A of the form

$$A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$$

where $A_{ii} : p_i \times p_i$ $i = 1,2$ is non-singular. The action on the sample space and parameter space are as before with $a \equiv 0$. A direct calculation shows that a maximal invariant parameter is $\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$ where

$$\delta_1 = \mu_1' \Sigma_{11}^{-1} \mu_1, \quad \delta_2 = (\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1)' \, \Sigma_{22\cdot1}^{-1} \, (\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1).$$

In terms of δ, the problem is to test that $\delta = 0$ versus $\delta \neq 0$. The analogue of Theorem 2.1 for the present problem will show there is no UMPI (under G_0) test of H_0 versus H_1.

<u>Theorem 2.2</u>: Let P_δ denote the probability measure of a maximal invariant at the parameter value δ. Then the Radon-Nikodym derivative dP_δ/dP_0 is given by

$$(2.4) \qquad R(t_1,t_2,t_3|\delta) = H(\delta)F_1(t_1|\delta_2)F_2|t_2|\delta_2)F_3(t_3|\delta_1)$$

where

$$H(\delta) = \exp\left[-\tfrac{1}{2}\frac{\delta_2}{c} - \tfrac{1}{2}\delta_1\left(\frac{1}{c}+\frac{1}{c_1}\right)\right]$$

$$(2.5) \quad \begin{cases} t_1 = \dfrac{Y_1' T_{11}^{-1} Y_1}{c} \\[2ex] t_2 = \left(\dfrac{Y_2}{\sqrt{c}} - \dfrac{T_{21}T_{11}^{-1}Y_1}{\sqrt{c}}\right)' T_{22\cdot1}^{-1}\left(\dfrac{Y_2}{\sqrt{c}} - \dfrac{T_{21}T_{11}^{-1}Y_1}{\sqrt{c}}\right) \\[2ex] t_3 = k\left(\dfrac{Y_1}{c} + \dfrac{X_1}{c_1}\right)'\left(V_1'V_1 + T_{11} + \dfrac{X_1X_1'}{c_1}\right)^{-1}\left(\dfrac{Y_1}{c} + \dfrac{X_1}{c_1}\right) \end{cases}$$

where $k^{-1} = \frac{1}{c} + \frac{1}{c_1}$ and

$$T \equiv V'V + \frac{1}{c}YY' = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}.$$

The functions F_i, $i = 1,2,3$ are given by

$$
(2.6)\quad \left\{\begin{aligned}
F_1(t_1|\delta_2) &= \exp\left[\tfrac{1}{2}\, t_1 \frac{\delta_2}{c}\right]\\
F_2(t_2|\delta_2) &= \sum_{j=0}^{\infty} d_j \left(t_2 \frac{\delta_2}{c}\right)^j\\
F_3(t_3|\delta_1) &= \sum_{j=0}^{\infty} h_j \left(t_3 \delta_1 \left(\frac{1}{c}+\frac{1}{c_1}\right)\right)^j
\end{aligned}\right.
$$

where the constants d_j and h_j are

$$
d_j = \frac{2^j}{(2j)!}\,\frac{\Gamma\left(\frac{p_2}{2}\right)}{\Gamma(\frac{1}{2})}\,\frac{\Gamma(j+\frac{1}{2})}{\Gamma\left(j+\frac{p_2}{2}\right)}\,\frac{\Gamma\left(\frac{n-p_1+1}{2}+j\right)}{\Gamma\left(\frac{n-p_1+1}{2}\right)}
$$

and

$$
h_j = \frac{2^j}{(2j)!}\,\frac{\Gamma\left(\frac{p_1}{2}\right)}{\Gamma(\frac{1}{2})}\,\frac{\Gamma(j+\frac{1}{2})}{\Gamma\left(j+\frac{p_1}{2}\right)}\,\frac{\Gamma\left(\frac{n+m_1+2}{2}+j\right)}{\Gamma\left(\frac{n+m_1+2}{2}\right)}
$$

Proof: A proof of this is similar to the proof of Theorem 3.1 which is outlined in Appendix II.

Expanding $R(t_1,t_2,t_3|\delta)$ about $\delta_1 = \delta_2 = 0$, the linear approximation to $R(t_1,t_2,t_3|\delta)$ is

$$
(2.7)\qquad R(t_1,t_2,t_3|\delta) = 1+\left(\frac{1}{c}+\frac{1}{c_1}\right)\left(h_1 t_3-\tfrac{1}{2}\right)\delta_1 + \left(\frac{d_1 t_2}{c} + \frac{t_1}{2c} - \frac{1}{2c}\right)\delta_2 + o(t_1,t_2,t_3,\delta)
$$

The remainder term is uniform in t_1,t_2 and t_3 since $0 \le t_i \le 1$ for $i = 1,2,3$. This implies that if ϕ is an invariant level α test of H_0 versus H_1, then the power function of ϕ, for small δ, is

$$
E_\delta\phi = \alpha+E_0\phi\{u_1\delta_1+u_2\delta_2\}+o(\delta)
$$

where

$$u_1 = \left(\frac{1}{c}+\frac{1}{c_1}\right)\left(h_1 t_3 - \tfrac{1}{2}\right)$$

$$u_2 = \left(\frac{d_1 t_2}{c} + \frac{t_1}{2c} - \frac{1}{2c}\right)$$

and the error term $o(\delta)$ is uniform in (t_1,t_2,t_3). Now consider testing $H_0:\delta_1=\delta_2=0$ versus $\tilde{H}_1:\delta_1=\gamma\delta_2>0$ where γ is a known positive constant. An easy application of the Generalized Neyman-Pearson Lemma shows that the level α test which rejects for large values of γu_1+u_2 is a LMPI test of H_0 versus $\tilde{H}_1$. Since this test depends on γ, there can be no LMPI test of H_0 versus H_1.

We now turn to a brief discussion of the likelihood ratio test of $H_0:\delta=0$ versus $H_1:\delta\neq 0$. A direct calculation shows that the LRT of $H_0^{(1)}:\mu_1=0$ rejects $H_0^{(1)}$ if

$$\lambda_1^{\frac{2}{m_1+n+2}} \equiv 1-t$$

is too small where t is defined by (2.3). Furthermore, the LRT of $H_0^{(2)}:\mu_2=0,\mu_1=0$ versus $H_1^{(2)}:\mu_2\neq 0,\mu_1=0$ rejects for small values of

$$\lambda_2^{\frac{2}{n+1}} \equiv \frac{\left|S_{22}-S_{21}S_{11}^{-1}S_{12}\right|}{\left|S_{22}+\frac{Y_2Y_2'}{c} - \left(S_{21}+\frac{Y_2Y_1'}{c}\right)\left(S_{11}+\frac{Y_1Y_1'}{c}\right)^{-1}\left(S_{21}+\frac{Y_2Y_1'}{c}\right)\right|}$$

where

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

is defined in Theorem 2.1. Now, the LRT of $H_0:\delta=0$ versus $H_1:\delta\neq 0$ rejects for small values of $\lambda_1\lambda_2$. In addition, under Ho, λ_1 and λ_2 are independent, $\lambda_1^{\frac{2}{m_1+n+2}}$ has a Beta distribution and $\lambda_2^{\frac{2}{n+1}}$ has a Beta distribution. (See Morrison and Bhoj (1973)). But, this does not yield the exact null distribution of $\lambda_1\lambda_2$ under H_0 expressed in terms of a tabled distribution. It should be mentioned that this type of decomposition of likelihood ratio statistics for testing normal means occurs in other contexts. For example, see Hogg (1961) for univariate normal example, Eaton (1972) for the MANOVA case, and Kariya (1974) for the application of these ideas to the multivariate linear growth curve model.

§3 Testing for Independence

In this section, we consider the problem of testing for independence based on data in canonical form. The canonical form will be of the type described by (1.1), but for simplicity our main discussion will be concerned with the following data: Consider three independent random matrices $V:n\times p, V_1:m_1\times p_1$ and $V_2:m_2\times p_2$ with $p_1+p_2 = p$ satisfying

$$
(3.1) \qquad \begin{aligned} L(V) &= N(0, I_n \otimes \Sigma) \\ L(V_i) &= N(0, I_{m_i} \otimes \Sigma_{ii}),\ i=1,2. \end{aligned}
$$

Here, the unknown covariance matrix Σ has been partitioned into $\Sigma_{ij}:p_i\times p_j$ for $i,j=1,2$. The data (3.1) arises from the data described in section one by assuming that the mean vector μ is known. The problem is to test $H_0:\Sigma_{12}=0$ versus the alternative $H_1:\Sigma_{12}\neq 0$. After describing our results for this problem, we will state some corresponding results for this testing problem when the data is given by (1.1) (and some minor variations of (1.1)).

The testing problem is invariant under the group G_2 whose elements g are

$$
g = A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{12} \end{pmatrix} \text{ with } A_{ii} \in G\ell_{p_i}, i=1,2.
$$

The action of g on a sample point is

$$g(V,V_1,V_2) = (VA',V_1A'_{11},V_2A'_{22})$$

with the action on Σ being

$$g(\Sigma) = A\Sigma A.$$

A maximal invariant parameter is the vector $\delta = (\delta_1^2,\cdots,\delta_q^2)'$ with $q = \min\{p_1,p_2\}$ where $\delta_1^2 \geq \cdots \geq \delta_q^2$ are the q-largest eigenvalues of $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$. The main result of this section is that there exists a LMPI test of the H_o versus H_1. To describe this result, let D_α^I be all the level α G_2-invariant test functions. Also, set

$$V'V = S = \begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22} \end{pmatrix}$$

and

$$V_i'V_i = W_{ii} \quad \text{for} \quad i=1,2.$$

<u>Theorem 3.1</u>: Let $\tau = \sum_1^q \delta_i^2$. For $\phi \in D_\alpha^I$, the power function of ϕ at δ, say $\pi(\phi,\delta)$, has the form

$$\pi(\phi,\delta) = \alpha + B(\phi)\,\tau + o(\delta,\phi) \tag{3.2}$$

where

$$\limsup_{\delta\to 0\;\; \phi} o(\delta,\phi) = 0$$

and

$$B(\phi) = E_o(\tfrac{1}{2}\phi\psi_o).$$

The statistic ψ_o is given by

$$(3.3) \qquad \psi_0 = \frac{(n+m_1)(n+m_2)}{p_1p_2} \operatorname{tr}(S_{11}+W_{11})^{-1}S_{12}(S_{22}+W_{22})^{-1}S_{21} + n - \sum_{i=1}^{2}\left(\frac{n+m_i}{p_i}\right)\operatorname{tr}(S_{ii}+W_{ii})^{-1}S_{ii}.$$

The level α test which rejects for $\psi_0 > k$ is a LMPI level α test.

Proof: The representation (3.2) is established in Appendix II. That rejecting for $\psi_0 > k$ gives a LMPI test follows immediately from (3.2) by maximimizing $B(\phi)$ and applying the generalized Neyman-Pearson Lemma.

In the discussion below, the situation treated by Theorem 3.1 for the data (3.1) will be called case (0). We now turn to a brief discussion of some other cases of interest.

case (i): This refers to case (0) when $m_1=0$ so that the data matrix V_1 is not available. A direct analogue of Theorem 3.1 shows that the test which rejects for large values of

$$(3.4) \qquad \psi_1 \equiv \frac{(n+m_2)n}{p_2p_1} \operatorname{tr} S_{11}^{-1}S_{12}(S_{22}+W_{22})^{-1}S_{21} - \frac{n+m_2}{p_2} \quad (S_{22}+W_{22})^{-1}S_{22}$$

is a LMPI test for testing $H_0:\Sigma_{12}=0$ versus $H_1:\Sigma_{12}\neq 0$.

case (ii): In this case we consider the data given in (1.1) and as usual, let $S = V'V, W_{ii} = V_i'V_i, i=1,2.$

Further, let

$$U = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \equiv \begin{pmatrix} (c+c_1)^{-\frac{1}{2}}(Y_1-X_1) \\ (c+c_2)^{-\frac{1}{2}}(Y_2-X_2) \end{pmatrix}$$

and set $b^2 = c^2/(c+c_1)(c+c_2)$. Define the statistic ψ_2 by

$$\begin{aligned}
\psi_2 = {} & \frac{(n+m_1+1)(n+m_2+1)}{p_1p_2} \operatorname{tr}(S_{11}+W_{11}+U_1U_1')^{-1}S_{12}(S_{22}+W_{22}+U_2U_2')^{-1}S_{21} \\
& +b^2\operatorname{tr}(S_{11}+W_{11}+U_1U_1')^{-1}U_1U_2'(S_{22}+W_{22}+U_1U_2')^{-1}U_2U_1' \\
& +n+b^2 - \frac{n+m_1+1}{p_1}\operatorname{tr}(S_{11}+W_{11}+U_1U_1')^{-1}(S_{11}+b^2U_1U_1') \\
& - \frac{n+m_2+1}{p_2}\operatorname{tr}(S_{22}+W_{22}+U_2U_2')^{-1}(S_{22}+b^2U_2U_2').
\end{aligned} \tag{3.5}$$

Rejecting $H_0:\Sigma_{12}=0$ for large values of ψ_2 is a LMPI test.

Case (iii): Again consider the data as given in (1.1) but assume that the mean of X_i is unrelated to the mean of Y, $i=1,2$. In this case, the LMPI test of $H_0:\Sigma_{12}=0$ rejects for large values of ψ_0 given by (3.3).

For the remainder of this section, we will be concerned only with the data given by (3.1), and the problem of testing $H_0:\Sigma_{12}=0$ versus $H_1:\Sigma_{12}\neq 0$. When both m_1 and m_2 are positive, we have been unable to calculate the likelihood ratio test (LRT) of H_0 versus H_1. However, in case (i) when $m_2=0$, the likelihood ratio is not difficult to derive.

Proposition 3.1: With the data of case (i), the LRT of H_0 versus H_1 rejects for small values of

$$\Lambda = |I-S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}| \tag{3.6}$$

Proof: This is a routine calculation and is omitted.

It is rather surprising that the LRT ignores the data V_1 in testing H_0. Of course, when $m_1=m_2=0$, rejecting for small Λ values gives the

LRT. However, when m_2 is very large, the value of Σ_{22} is essentially known but the likelihood ratio criterion ignores this information. Indeed, if Σ_{22} is known, the LRT, based on S alone, for testing H_0: $\Sigma_{12} = 0$ is also that given in Proposition 3.1.

It is not clear what to do in practice for testing H_0 versus H_1. One possibility is to ignore the two ancillary statistics $\mathrm{tr}(S_{ii}+W_{ii})^{-1}S_{ii}$, $i=1,2$, in (3.3) and reject H_0 for large values of

$$\tilde{\psi}_1 = \mathrm{tr}(S_{11}+W_{11})^{-1}S_{12}(S_{22}+W_{22})^{-1}S_{21}.$$

The null distribution of ψ_0 and $\tilde{\psi}_1$ is not known. Letting n and m_i tend to ∞ with $\frac{m_i}{n} \to \beta_i$, $i=1,2$, it is not too hard to show that $n\tilde{\psi}_1$ converges in distribution to a random variable with a scaled chi-square distribution. In particular,

$$n\tilde{\psi}_1 \xrightarrow{d} (1+\beta_1)(1+\beta_2)\chi^2_{p_1 p_2}.$$

When n is large, this provides one possible method of testing H_0 versus H_1.

Now, consider the special case when $p_2=1$ and $m_1=0$. The problem of testing for independence is similar in structure to a mean testing problem discussed by Giri (1968). Even though the maximal invariant parameter is one dimensional, a uniformly most powerful invariant test does not exist and the LRT is not the locally best test (Theorem 3.1). As in the problem treated by Giri (1968), a natural maximal invariant in the sample space is two dimensional say (ξ_1,ξ_2), and ξ_2 is an ancillary statistic. The LRT rejects for large values of ξ_1 while the locally best test involves both ξ_1 and ξ_2. The details of this are given in Eaton and Kariya (1974). A related reference is Marden (1978).

Finally, consider the special case of $p_1 = p_2 = 1$ so $p=2$ and the data is given by 3.1. A minimal sufficient statistic is (S,W_{11},W_{22}) where

$$S = V'V, W_{ii} = V_i'V_i, \ i=1,2.$$

In this case, the problem is to test that $\rho=0$ where ρ is the bivariate correlation coefficient. The testing problem is invariant under scale changes and a maximal invariant statistic is $T = (t_1,t_2,t_3)$ where

$$t_1 = \frac{S_{12}^2}{S_{11}S_{22}},\ t_2 = \frac{W_{11}}{S_{11}},\ t_3 = \frac{W_{22}}{S_{22}}.$$

When $m_1=0$ (so W_{11} is not present and t_2 is not present), the LRT rejects for large values of t_1 while the LMPI test involves both t_1 and t_3. Since t_3 is ancilliary, it may be most reasonable to condition on t_3 and test $\rho=0$ conditionally. But, when both m_1 and m_2 are positive there is a complication. The statistics t_2 and t_3 are marginally ancillary but (t_2,t_3) is not an ancillary statistic. It is not clear how to condition in this case, but rejecting for large t_1 is not appropriate.

§ 4: Examples

In this section, we present two problems which are special cases of the problems discussed in section 3. The notation used in these two examples is independent of the notation used earlier.

Example 4.1: In this example, we consider the problem of testing independence in a model of covariate discriminant analysis. Suppose $X_1,\dots,X_M$ are i.i.d. $N_p(\mu,\Sigma)$ and $Y_1,\dots,Y_N$ are i.i.d. $N_p(\nu,\Sigma)$ and write the dimension parameter p as $p = p_1+p_2$. Partitioning the data and parameters, we have

$$X_j = \begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix}, \quad Y_j = \begin{pmatrix} Y_j^{(1)} \\ Y_j^{(2)} \end{pmatrix},$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \nu = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} .$$

Of course, $X_i^{(\alpha)}$, and $Y_j^{(\alpha)}$, μ_α and ν_α are $p_\alpha \times 1$ and $\Sigma_{\alpha\beta}$ is $p_\alpha \times p_\beta$, $\alpha,\beta = 1,2$. It is assumed that $\mu_2 = \nu_2$. Discrimination problems in this situation have been considered by Cochran and Bliss (1948), Rao (1949), Cochran (1964), Rao (1966) and Memon and Okamoto (1970). A survey of this situation is given in Kshirsargar (1972, p. 200-203). When Σ is known and $\Sigma_{12} \neq 0$, Cochran and Bliss (1948) constructed a discriminant function based on all the data which is more efficient than the usual discriminant function based on $X_i^{(1)}$, $i = 1, \dots, M$ and $Y_j^{(1)}$, $j = 1, \dots, N$. When Σ is unknown, Cochran and Bliss (1948) proposed a discriminant function in which Σ is replaced by an estimate. However, when Σ_{12} is close to zero, this discriminant function does not seem to be better than the usual one based on $X_i^{(1)}$ and $Y_j^{(1)}$, $i = 1, \dots, M$, $j \equiv 1, \dots, N$. Of course, when $\Sigma_{12} = 0$, it seems most reasonable to base discrimination solely on the basis of $X_i^{(1)}$ and $Y_j^{(1)}$,

$i = 1, \ldots, M$ and $j = 1, \ldots, N$. This motivates the problem of testing $\Sigma_{12} = 0$ in this situation.

After a reduction by invariance, we will show that testing $\Sigma_{12} = 0$ is a special case of the problem described in section 3. As demonstrated in section 1, the data $X_1, \ldots, X_M$ is equivalent (via a linear transformation) to (U_1, V_1) where $U_1 : p \times 1$ and $V_1 : (M-1) \times p$ are independent and

$$L(U_1) = N(\mu, \frac{1}{M}\Sigma), \; L(V_1) = N(0, I_{M-1} \otimes \Sigma) \quad .$$

Similarly $Y_1, \ldots, Y_N$ is equivalent to (U_2, V_2) which are independent and

$$L(U_2) = N(\nu, \frac{1}{N}\Sigma), \; L(V_2) = N(0, I_{N-1} \otimes \Sigma) \quad .$$

The problem of testing $\Sigma_{12} = 0$ is obviously invariant under translations of U_1 and U_2 given by

$$U_1 \to U_1 + \binom{a}{c} \; , \; U_2 \to U_2 + \binom{b}{c}$$

with $a, b \in R^{p_1}$ and $c \in R^{p_2}$. A maximal invariant is $W_2 \equiv k(U_1^{(2)} - U_2^{(2)})$, $k = (\frac{1}{M} + \frac{1}{N})^{-\frac{1}{2}}$. After this reduction by invariance, the data is W_2 and $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$ where

$$L(W_2) = N(0, \Sigma_{22}), \; L(V) = N(0, I_{M+N-2} \otimes \Sigma)$$

and the problem is to test $\Sigma_{12} = 0$. In this form, the results of Theorem 3.1 are applicable (with $m_1 = 0$ and $W_{11} = 0$), so a locally most powerful invariant test exists and is given in Theorem 3.1. It is interesting to note that the LRT based on all the data of $H_0 : \Sigma_{12} = 0$ versus $H_1 : \Sigma_{12} \neq 0$ is the same as the LRT based only on the data matrix V. In other words, the LRT described in Proposition 3.1 ignores the extra information that

$\mu_2 = \nu_2$. However, the LRT would not ignore the information that $\mu = \nu$.

Our second example concerns the growth curve model.

Example 4.2: Consider a data matrix $Y:N\times p$ such that

$$L(Y) = N(X_1BX_2, I_N\otimes\Omega)$$

where X_1 is $N\times r$ of rank r, X_2 is $q\times p$ of rank q , and both are known (see Potthoff and Roy (1964)). Also, $B:r\times q$ is a matrix of unknown parameters and Ω is a $p\times p$ positive definite matrix. The estimation of B is of concern here and of course the structure of Ω affects the estimation. Let Z_2 be a $(p-q)\times p$ matrix of rank $p-q$ such that $X_2Z_2' = 0$. If Ω has the form

$$\Omega = X_2'\psi_1X_2 + Z_2'\psi_2Z_2 \tag{4.1}$$

where $\psi_1:q\times q$ is positive definite and $\psi_2:(p-q)\times p-q$ is positive definite, then the least squares (acting as if $\Omega = I_p$) estimator of B is also the Gauss-Markov and maximum likelihood estimator of B. This claim and its converse follow from results in Eaton (1970), or a modification of a result due to Rao (1967). The covariance structure (4.1) is known as Rao's covariance structure (Rao(1967)) and has been discussed in Geisser (1970). Further Lee and Geisser (1972) derived the LRT for testing that Ω has the form (4.1) versus arbitrary alternatives. We will show that, after a reduction by invariance, testing Ω has the form (4.1) is a special case of testing for independence with additional information. First, a transformation to a canonical form will simplify certain calculations. Let $Z_1:N\times(N-r)$ be of rank $(N-r)$ and satisfy $Z_1'X_1 = 0$. Then, let

$$\Gamma_1 = [X_1(X_1'X_1)^{-\frac{1}{2}}, Z_1(Z_1'Z_1)^{-\frac{1}{2}}] ,$$

$$\Gamma_2 = [X_2'(X_2X_2')^{-} \ , \ Z_2'(Z_2Z_2')^{-\frac{1}{2}}]$$

so Γ_1 is $N\times N$ and orthogonal and Γ_2 is $p\times p$ and orthogonal. Now, let

$$W = \Gamma_1'Y\Gamma_2, \quad \Sigma = \Gamma_2'\Omega\Gamma_2, \quad \mu = (X_1'X_1)^{\frac{1}{2}}B(X_2X_2')^{\frac{1}{2}}$$

and partition W and Σ as

$$W = \begin{array}{cc} \begin{array}{cc} q & p-q \end{array} & \\ \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} & \begin{array}{c} r \\ N-r \end{array} \end{array} \quad , \quad \Sigma = \begin{array}{cc} \begin{array}{cc} q & p-q \end{array} & \\ \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} & \begin{array}{c} q \\ p-q \end{array} \end{array} \quad .$$

With this relabeling, we have

$$L(W) = N\left(\begin{pmatrix} \mu & 0 \\ 0 & 0 \end{pmatrix} , I_N \otimes \Sigma\right),$$

The null hypothesis that Ω has the form (4.1) becomes $H_0:\Sigma_{12} = 0$ when expressed in terms of Σ. This testing problem is invariant under the translations $W_{11} \to W_{11}+a$, $a:r\times q$ and a maximal invariant under this group of translations is $\{W_{12}, (W_{21}, W_{22})\}$. Clearly W_{12} is independent of (W_{21}, W_{22}) and

$$L(W_{12}) = N(0, I_r \otimes \Sigma_{22})$$

$$L((W_{21}, W_{22})) = N(0, I_{N-r} \otimes \Sigma) .$$

Based on this data, testing H_0 is a special case of the problem treated in case (i) following Theorem 3.1. Also, the result described in Proposition 3.1 shows that the LRT based on $\{W_{12}, (W_{21}, W_{22})\}$ ignores W_{12} and the LRT is different from the locally best test. Furthermore, this LRT is the same as the LRT based on all the data W which Lee and Geisser (1972) derived.

Appendix I

In this appendix we will describe a recent result due to S. Andersson (1979) concerning quotient measures and the representation of densities of maximal invariants. Because of the simplicity of Andersson's conditions, we have found this result easier to apply than a similar representation due to Wijsman (1967). In what follows the notation and terminology given in Nachbin (1965) will be used. Let G be a locally compact σ-compact topological group which acts topologically on the locally compact σ-compact space X. A left (right) Haar measure on G is denoted by $\nu_\ell(\nu_r)$ and the right hand modulus of G is Δ_r so $\Delta_r\nu_r = \nu_\ell$. The natural projection π from X to the quotient space $X/G = Q$ is a convenient and natural choice for a maximal invariant under the action of G on X. All measures on X and Q will be Radon measures.

Let μ be a measure on X which is relatively invariant with multiplier Δ_r^{-1} -- that is,

$$\int_X f(g^{-1}x)\mu(dx) = \Delta_r^{-1}(g) \int_X f(x)\mu(dx)$$

for all μ-integrable f. Ignoring questions of existence of integrals for a moment, consider

$$\tilde{f}(x) \equiv \int_G f(gx)\nu_r(dg) \tag{A.1}$$

and note that $\tilde{f}(x) = \tilde{f}(hx)$, $h \in G$ so $\tilde{f}$ is invariant. Thus, we can write $\tilde{f}(x) = \hat{f}(\pi(x))$, where $\hat{f}$ is defined on Q. If α is a measure on Q, we can then integrate $\hat{f}$ over Q. This integration will be denoted by

$$J_\alpha(f) \equiv \int_Q\left(\int_G f(gx)\ \nu_r(dg)\right) d\alpha. \tag{A.2}$$

An easy calculation shows that (A.2) is relatively left invariant with multiplier Δ_r^{-1} -- that is, $J_\alpha(g_1 f) = \Delta_r^{-1}(g_1)\, J_\alpha(f)$. A question treated in Andersson (1978) is the following -- given μ on X which is relatively invariant with multiplier Δ_r^{-1} , under what conditions will there exist an α on Q so that

$$(A.3) \qquad J_\alpha(f) = \int_X f(x)\, \mu(dx)$$

for all μ-integrable f? A sufficient condition for the representation (A.3) to hold is provided by the notion of a proper action (see Andersson (1978)).

Definition: Consider the mapping K of $G\times X$ to $X\times X$ given by $K(g,x) = (gx,x)$. The action of G on X is proper if $K^{-1}(C)$ is compact for each compact subset $C \subseteq X\times X$.

Theorem (see Andersson (1978)). Suppose the action of G on X is proper. If μ is a relatively left invariant measure on X with multiplier Δ_r^{-1}, then there exists a measure α on Q such that

$$(A.4) \qquad \int_X f(x)\, \mu(dx) = \int_Q \left(\int_G f(gx)\, \nu_r(dg) \right) d\alpha$$

for all μ-integrable f.

For the remainder of this appendix, it is assumed that G acts properly on X. In some situations, one has a measure μ_0 on X which is relatively invariant with a multiplier χ_0 , -- that is

$$\int f(g^{-1}x)\, \mu_0(dx) = \chi_0(g) \int f(x)\, \mu_0(dx)$$

for all integrable f and $g \in G$. But a representation of the form (A.4) is still desired. To obtain such a result, the measure μ_0 needs to be modified. It is asserted in Andersson (1979) that there exists a positive continuous function η_0 on X such that

(A.5) $\quad \eta_0(gx) = \Delta_r(g)\, \chi_0(g)\, \eta_0(x)$

for $x \in X$ and $g \in G$. Setting $\mu = \eta_0^{-1}\mu_0$, it is easily verified that μ is relatively invariant with multiplier Δ_r^{-1}. Thus, applying the above theorem gives

(A.6) $\quad \int_X f(x)\, \mu_0(dx) = \int_X f(x)\, \eta_0(x)\, \mu(dx) =$

$$\int_Q \left(\int_G f(gx)\, \eta_0(gx)\, \nu_r(dg) \right) d\alpha =$$

$$\int_Q \left(\eta_0(x) \int_G f(gx)\, \chi_0(g)\, \nu_\ell(dg) \right) d\alpha \ .$$

where $\nu_\ell = \Delta_r \nu_r$ is a left Haar measure on G.

Following Andersson (1979), we will now apply (A.6) to find the density function of a maximal invariant. Let μ_0 be relatively invariant with multiplier χ_0 and suppose that the random variable $X \in X$ has a density f_0 with respect to μ_0. The random variable $Y \equiv \pi(X) \in Q$ is maximal invariant. In the notation of (A.6), the claim is that the density of Y with respect to the measure α is φ_0 where

(A.7) $\quad \varphi_0(\pi(x)) = \eta_0(x) \int_G f_0(gx)\, \chi_0(g)\, \nu_\ell(dg)\ .$

Given (A.6), the verification of (A.7) is identical to the case when G is compact (see Stein (1966) or Eaton (1972)). To verify that φ_0 is the density of Y, it suffices to show that

$$\mathcal{E}k(Y) = \int_Q k(y)\, \varphi_0(y)\, \alpha(dy)$$

for suitably many functions k on Q. But,

$$\mathcal{E}k(Y) = \mathcal{E}k(\pi(X)) = \int k(\pi(x))\, f_0(x)\, \mu_0(dx) =$$

$$\int_Q \eta_0(x)\ k(\pi(x)) \int_G f_0(gx)\ \chi_0(g)\ \nu_\ell(dg)\quad d\alpha =$$

$$\int_Q k(y)\ \varphi_0(y)\ \alpha(dy)$$

with the last equality following from the definition of φ_0 .

Our application of (A.7) concerns the ratio of two densities of a maximal invariant. If f_0 and f_1 are two possible densities of X and φ_0 and φ_1 are the two induced densities of Y, then from (A.7) the ratio $r(y) = \varphi_1(y)/\varphi_0(y)$ is given by

$$r(y) = \frac{\int_G f_1(gx)\ \chi_0(g)\ \nu_\ell(dg)}{\int_G f_0(gx)\ \chi_0(g)\ \nu_\ell(dg)} \tag{A.8}$$

as long as the denominator is positive.

Appendix II

The purpose of this appendix is to outline proofs of Theorems 2.1, 2.2 and Theorem 3.1. We will restrict our attention to the proof of Theorem 3.1. The other Theorems are proved using similar ideas.

Of course, the basic idea in the proof of Theorem 3.1 is to use equation (A.8) to express the density of a maximal invariant for small parameter values (see Giri and Kiefer (1964) or Schwartz (1967)). The natural sample space of the data given by (3.1) is $\mathcal{X} = \mathcal{X}_0 \times \mathcal{X}_1 \times \mathcal{X}_2$ where $\mathcal{X}_0$ is the linear space of all $n \times p$ real matrices and $\mathcal{X}_i$ is the linear space of all $m_i \times p_i$ real matrices for $i = 1, 2$. A point $x \in \mathcal{X}$ will sometimes be written $x = (x_0, x_1, x_2)$ with $x_i \in \mathcal{X}_i$, $i = 1, 2, 3$. Also, dx (or dx_i) will denote Lebesque measure on $\mathcal{X}(\mathcal{X}_i)$. In the notation of (3.1), the density of $X = (V, V_1, V_2) \in \mathcal{X}$ is

$$f(x|\Sigma) = \prod_{i=1}^{2} \frac{|\Sigma_{ii}|^{-\frac{m_i}{2}}}{(\sqrt{2\pi})^{m_i}} \exp[-\tfrac{1}{2} \operatorname{tr} x_i \Sigma_{ii}^{-1} x_i'] \frac{|\Sigma|^{-\frac{n}{2}}}{(\sqrt{2\pi})^{n}} \exp[-\tfrac{1}{2} \operatorname{tr} x_0 \Sigma^{-1} x_0']$$

The group G_1 defined in section 3 acts on a point $x \in \mathcal{X}$ by

$$A(x) = (x_0 A', x_1 A_1', x_2 A_2')$$

where

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad A_i \in G\ell_{p_i}, \ i = 1, 2 .$$

It is easily verified that

$$\nu_\ell(dA) = \frac{dA_1}{|A_1 A_1'|^{p_1/2}} \frac{dA_2}{|A_2 A_2'|^{p_2/2}}$$

is a left invariant measure on G_1. Further, dx is relatively invariant with multiplier

$$\chi_0(A) = \prod_{i=1}^{2} |A_i A_i'|^{(n+m_i)/2} .$$

To establish Theorem 3.1, we will use the following well known argument. For any invariant test function φ, the power function of φ at a maximal invariant parameter point δ is

$$\text{(B.0)} \qquad \pi(\varphi,\delta) = \int \varphi\, dP_\delta = \int \varphi \left(\frac{dP_\delta}{dP_0}\right) dP_0$$

where P_δ is the probability measure of a maximal invariant. However, under the conditions given in Appendix I, the ratio dP_δ/dP_0 is given by

$$\text{(B.1)} \qquad r_\delta(x) = \frac{\int_{G_1} f(Ax|\Sigma(\delta))\ \chi_0(A)\ \nu_\ell(dA)}{\int_{G_1} f(A(x)|\Sigma(0))\,\chi(A)\ \nu_\ell(dA)}$$

where

$$\Sigma(\delta) \equiv \begin{pmatrix} I_{p_i} & \Delta \\ \Delta' & I_{p_2} \end{pmatrix}$$

and Δ is $p_1 \times p_2$ with $\Delta_{ii} = \delta_{ii}$ for $i = 1, \ldots, p_2$, and $\Delta_{ij} = 0$ for $i \neq j$. Without loss of generality we have taken $p_1 \geq p_2$. Before calculating r_δ, the assumptions required to apply (A.8) need to be checked. That X and G_2 are locally compact sigma compact spaces is clear and obviously G_2 acts topologically on X. Thus, it must be shown that the action of G_2 on X is proper. The next paragraph is devoted to a discussion of this point.

First, the action of G_2 on X is not proper. To see this, take $C = \{0,0\} \in X \times X$ so

$$K^{-1}(C) = G_1 \times \{0\}$$

which is not compact in $G_2 \times X$. The work of Wijsman (1967) and the discussion in Farrell (1976) suggests that a set of Lebesque measure zero needs to be removed from X before (A.8) can be applied. For $x = (x_0, x_1, x_2)$, write

$$y_1 = \begin{pmatrix} x_0^{(1)} \\ x_1 \end{pmatrix}, \quad y_2 = \begin{pmatrix} x_0^{(2)} \\ x_2 \end{pmatrix}$$

where y_i is $(n+m_i) \times p_i$, $i = 1, 2$ and x_0 has been partitioned into $x_0^{(i)}: n \times p_i$, $i = 1, 2$. Now, the action of G_2 on X becomes

$$A(y_1, y_2) \equiv (y_1 A_1', y_2 A_2')$$

where

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \in G_2 \ .$$

Let Y_i be the real linear space of all $(n+m_i) \times p_i$ matrices so $G_2 = G\ell_{p_1} \times G\ell_{p_2}$ acts on $Y_1 \times Y_2 \equiv Y$ coordinatewise. Furthermore, Lebesque measure dx on X corresponds to Lebesque measure dy on Y. Since the product group $G\ell_{p_1} \times G\ell_{p_2}$ acts coordinatewise on the product space $Y_1 \times Y_2$ and $dy = dy_1 \times dy_2$, to discuss the issues surrounding the action of G_2 on Y, it suffices to discuss the action of $G\ell_{p_1}$ on Y_1. Let N_1 be the elements of Y_1 which have rank less than p_1. Since $n+m_1 \geq p_1$, N_1 has Lebesque measure zero and the space $Y_1^* \equiv Y_1 - N_1$ is an open set in Y_1 which is acted on by $G\ell_{p_1}$. The claim is that $G\ell_{p_1}$ acts properly on Y_1^* . To see this, let $C \subseteq Y_1^* \times Y_1^*$ be compact. It is clear that we can find a compact set $B_1 \subseteq Y_1^*$ such that $C \subseteq B_1 \times B_1$. Since the mapping K defined in Appendix I is continuous, it follows that if $K^{-1}(B_1 \times B_1)$ is compact then $K^{-1}(C)$ is compact. However,

$$K^{-1}(B_1 \times B_1) = \{(g,y) | (g(y), y) \in B_1 \times B_1\} \ .$$

For $u \in Y_1$, let

$$\|u\|^2 = \text{tr } u'u \equiv \text{trace } (u'u)$$

so $\|\cdot\|$ is a norm on $\mathcal{Y}_1$. Also for $h \in R^{p_1^2}$, let

$$\|h\|_1^2 = \text{tr } hh'$$

so $\|\cdot\|_1$ is a norm on $R^{p_1^2}$. Since $K^{-1}(B_1 \times B_1)$ is a closed subset of $G\ell_{p_1} \times \mathcal{Y}_1^* \subseteq R^{p_1^2} \times \mathcal{Y}_1$, to show $K^{-1}(B_1 \times B_1)$ is compact, it suffices to show that $\|g\|_1^2 + \|y\|^2$ remains bounded for $(g,y) \in K^{-1}(B_1 \times B_1)$. The compactness of $B_1 \subseteq \mathcal{Y}_1^*$ implies that there are two constants a_1 and b_1 such that

$$\|y\|^2 \leq a_1 < +\infty \text{ , } y \in B_1$$

and

$$\lambda_p(y'y) \geq b_1 > 0 \text{ , } y \in B_1$$

where $\lambda_p(y'y)$ is the smallest eigenvalue of $y'y$. For $(g,y) \in K^{-1}(B_1 \times B_1)$,

$$a_1 \geq \|g(y)\|^2 = \text{tr}[(yg')'yg'] =$$

$$\text{tr } y'y\, g'g = \text{tr}(y'y - b_1 I_{p_1})g'g + b_1 \text{ tr } g'g \geq b_1 \|g\|_1^2 \text{ .}$$

The inequality follows since for $y \in B_1$, $y'y - b_1 I_p$ is positive definite so $\text{tr}(y'y - b_1 I_{p_1})g'g \geq 0$. Therefore, for $(g,y) \in K^{-1}(B_1 \times B_1)$,

$$\|g_1\|^2 + \|y\|^2 \leq a_1 + a_1 b_1^{-1}$$

so $K^{-1}(B_1 \times B_1)$ is compact. Hence the action of $G\ell_{p_1}$ on $\mathcal{Y}_1^*$ is proper. Now, let $X^* \subseteq X$ be defined by

$$X^* = \{ (x_0, x_1, x_0) | \begin{pmatrix} x_0^{(i)} \\ x_i \end{pmatrix} \text{ has rank } p_i, \ i = 1,2\} \text{ .}$$

It follows immediately that $X - X^*$ has Lebesque measure zero and G_2 acts properly on X^*.

We now proceed with the evaluation of $r_\delta(x)$ given in (B.1). For $x = (x_0, x_1, x_2) \in X^*$, let

$$x_0'x_0 \equiv S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} ,$$

$$x_i'x_i \equiv W_{ii} , \quad i = 1, 2$$

and

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \equiv \begin{pmatrix} (S_{11}+W_{11})^{-\frac{1}{2}} & 0 \\ 0 & (S_{22}+W_{22})^{-\frac{1}{2}} \end{pmatrix} S \begin{pmatrix} (S_{11}+W_{11})^{-\frac{1}{2}} & 0 \\ 0 & (S_{22}+W_{22})^{-\frac{1}{2}} \end{pmatrix}$$

where the notation of §3 is being used. Note that $0 < T < I_p$ in the sense of positive definiteness. Let $\mathcal{E}_{k,\alpha}$ denote expectation with respect to the distribution on $G\ell_k$ given by

$$P(dg) = c(\alpha,k)\,|gg'|^{\alpha/2}\exp[-\tfrac{1}{2}\operatorname{tr} gg']\ \nu_\ell(dg)$$

where $\alpha > 0$, $c(\alpha,k)$ is a normalization constant and

$$\nu_\ell(dg) = \frac{dg}{|g'g|^{k/2}} .$$

A bit of algebra and a change of variable show that

$$\text{(B.2)} \qquad r_\delta(x) = |\Sigma(\delta)|^{-n/2}\mathcal{E}_{p_1,n+m_1}\ \mathcal{E}_{p_2,n+m_2}\{\exp[-\tfrac{1}{2}\operatorname{tr} TA'(\Sigma^{-1}(\delta)-I_p)A]\}$$

where

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} , \quad A_i \in G\ell_{p_i} , \quad i = 1, 2,$$

and $\mathcal{E}_{p_i,n+m_i}$ is expectation on A_i, $i = 1, 2$. Define $\gamma: p\times p$ by

$$\gamma \equiv \Sigma^{-1}(\delta)-I_p = \begin{pmatrix} (I_1-\Delta\Delta')^{-1}-I_1 & -(I_1-\Delta\Delta')^{-1}\Delta \\ -(I_2-\Delta'\Delta)^{-1}\Delta' & (I_2-\Delta'\Delta)^{-1}-I_2 \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

where γ_{ij} is $p_i \times p_j$, i = 1, 2. Then we have

$$\mathrm{tr}\ TA'(\Sigma^{-1}(\delta)-I_p)A = \mathrm{tr}\ T_{11}A_1'\gamma_{11}A_1 + 2\mathrm{tr}\ T_{12}A_2'\gamma_{21}A_1 + \mathrm{tr}\ T_{22}A_2'\gamma_{22}\ A_2 \ .$$

We now make the following claim: For δ small, $\tau = \Sigma_1^{p_2}\ \delta_i^2$, and all T satisfying $0 < T < I_p$ (in the sense of positive definiteness),

$$\text{(B.3)}\quad \begin{cases} \exp[-\tfrac{1}{2}\mathrm{tr}T_{11}A_1'\gamma_{11}A_1] = 1 - \tfrac{1}{2}\mathrm{tr}\ T_{11}A_1'\Delta\Delta' A_1 + R_1(T_{11},\ A_1,\ \Delta) \\ \exp[-\tfrac{1}{2}\mathrm{tr}T_{22}A_2'\gamma_{22}A_2] = 1 - \tfrac{1}{2}\mathrm{tr}\ T_{22}A_2'\Delta'\Delta A_2 + R_2(T_{22},\ A_2,\ \Delta) \\ \exp[-\mathrm{tr}\ T_{12}A_2'\gamma_{21}A_1] = 1 + \mathrm{tr}\ T_{12}A_2'\Delta' A_1 + \tfrac{1}{2}(\mathrm{tr}T_{12}A_2'\Delta' A_1)^2 + \\ \qquad\qquad R_3(T_{12},\ A_1,\ A_2,\ \Delta) \end{cases}$$

where the error terms R_1, R_2 and R_3 satisfy

$$\text{(B.4)}\quad \begin{cases} |R_1(T_{11},\ A_1,\ \Delta| \le H_1(A_1)\ \varphi_1(\tau) \\ |T_2(T_{22},\ A_2,\ \Delta)| \le H_2(A_2)\ \varphi_2(\tau) \\ |R_3(T_{12},\ A_1,\ A_2,\ \Delta) \ \le H_3(A_1)\ H_4(A_2)\ \varphi_3(\tau) \end{cases}$$

Further, the inequalities in (B.4) hold for all T, $0 < T < I_p$, the functions H_i are integrable (A_1, A_2) and

$$\lim_{\tau\to 0} \varphi_i(\tau)/\tau = 0 \ ,\ i = 1,\ 2,\ 3 \ .$$

The arguments leading to (B.3) and (B.4) are similar to those in Schwarz(1967) and Kariya (1978) and are omitted. The following identities are used in the evaluation of (B.2):

$$\text{(B.5)}\quad \begin{cases} E_{p_1, n+m_i}\ \mathrm{tr}\ T_{11}A_1'\Delta\Delta' A_1 = \dfrac{n+m_1}{p_1}\ (\mathrm{tr}\ T_{11}T_{11}')\tau \\ E_{p_2, n+m_2}\ \mathrm{tr}\ T_{22}'A_2\Delta'\Delta A_2 = \dfrac{n+m_1}{p_2}\ (\mathrm{tr}\ T_{22}T_{22}')\ \tau \end{cases}$$

$$(B.5)\quad \begin{cases} E_{p_1,n+m_1} E_{p_2,n+m_2} \operatorname{tr} T_{12}A_2'\Delta' A_1 = 0 \\ E_{p_1,n+m_1} E_{p_2,n+m_2} (\operatorname{tr} T_{12}A_2'\Delta' A_1)^2 = \dfrac{n+m_1}{p_1}\dfrac{n+m_2}{p_2}(\operatorname{tr} T_{12}T_{12}')\tau \end{cases}$$

Note that $|\Sigma(\delta)|^{-n/2} = 1 + \frac{n}{2}\tau + o(\delta)$ where $\lim_{\delta\to 0} o(\delta)/\tau = 0$. Substituting this and the expressions in (B.3) into (B.2) leads to the expression:

$$(B.6)\qquad r_\delta(x) = 1 + \frac{1}{2}\psi_0\tau + o(T,\delta)$$

where ψ_0 is defined in (3.3). The remainder term is uniformly bounded in T, $0 < T < I_p$, and satisfies

$$\lim_{\delta\to 0}\ \sup_{0<T<I_p} \frac{o(T,\delta)}{\tau} = 0 \quad .$$

The identities in (B.5) and the results expressed in (B.4) are used to establish (B.6).

Now, let φ be any level α invariant test of $H_0:\delta = 0$ versus $H_1:\delta \neq 0$. Substituting (B.6) into (B.0) yields

$$\pi(\varphi,\delta) = \int\varphi\left(\frac{dP_\delta}{dP_0}\right)dP_0 = \int\varphi[1 + \frac{1}{2}\psi_0\tau + o(T,\delta)]dP_0 =$$

$$\alpha + \frac{1}{2}(E_0\varphi\psi_0)\tau + o(\varphi,\delta)$$

where the remainder term satisfies

$$\lim_{\delta\to 0}\ \sup_{\varphi} \frac{o(\varphi,\delta)}{\tau} = 0 \ .$$

This proves Theorem 3.1.

References

[1] Andersson, S.A. (1978). Invariant Measures. Stanford University Technical Report No. 129.

[2] Andersson, S.A. (1979). Distribution of maximal invariants using proper action and quotient measures. Preprint No. 2, 1979. Institutue of Mathematical Statistics, University of Copenhagen.

[3] Bhargava, R.N.P. (1962). Multivariate tests of hypotheses with incomplete data. Technical Report No. 3, Applied Mathematics and Statistical Laboratories, Stanford University.

[4] Cochran, W.G. and Bliss, C.I. (1948). Discriminant functions with covariance. Ann. Math. Statist. 19, p. 151.

[5] Cochran, W.G. (1964). Comparison of two methods of handling covariates in discriminatory analysis. Ann. Inst. Statist. Math. 16, p. 43.

[6] Eaton, M.L. (1970). Gauss-Markov estimation for multivariate linear models: A coordinate free approach. Ann. Math Statist. 44, p 528-538.

[7] Eaton, M.L. (1972). Multivariate Statistical Analysis. Institute of Math. Statist., University of Copenhagen, Copenhagen, Denmark.

[8] Eaton, M.L. and Kariya, T. (1974). "Testing for independence with additional information", University of Minnesota Technical Report No. 238.

[9] Eaton, M.L. and Kariya, T. (1974). Tests and means with additional information. University of Minnesota Technical Report No. 243.

[10] Ekbohm, G. (1976). On comparing means in the paired case with incomplete data on both responses. Biometrika, 63, p.299-304.

[11] Farrell, R. (1976). Techniques of Multivariate Calculation. Springer Verlag. New York.

[12] Geisser, S. (1970). Bayesian Analysis of Growth Curves. Sankhya, Series A, 32, p. 53-64.

[13] Giri, N. (1968). Locally and asymptotic tests of a multivariate problem. Ann. Math. Statist. 39, p. 171-178.

[14] Giri, N. and Kiefer, J. (1964). Local and asymptotic minimax properties of a normal multivariate testing problem. Ann. Math. Statist., 35. p. 21-35.

[15] Hartley, H.O. and Hocking, R.R. (1971). The analysis of incomplete data. Biometrics, 27, p. 783-823.

[16] Hogg, R.V. (1961). "On the resolution of statistical hypotheses". Jour. Amer. Statist. Ass'n., 56, p. 978-89.

[17] Kariya, T. (1974). Hypothesis testing problems in the general MANOVA model. Ph.D. Thesis. University of Minnesota.

[18] Kariya, T. (1978). The general MANOVA problem. Ann. Statist. 6, p. 200-214.

[19] Kshirsargar, A.M. (1972). Multivariate Analysis. Marcel Decker, New York, N.Y.

[20] Lee, J.C. and Geisser, S. (1972). Growth Curve Prediction. Sankhyā, Series A, 34, p. 393-412.

[21] Little, R.J.A. (1976). Inference about means from incomplete multivariate data. Biometrika, 63, p. 593-604.

[22] Marden, J. (1978). Invariant tests on covariance matrices. Submitted for publication.

[23] Memon, A.Z. and Okamoto, M. (1970). The classification statistic W^* in covariate discriminant analysis. Ann. Math Statist. 41, p. 1491-9.

[24] Morrison, D.F. and Bhoj, D.S. (1973). Power of the likelihood ratio test on the mean vector of the multivariate normal distribution with missing observations. Biometrika, 60, p. 365-368.

[25] Nachbin, L. (1965). The Haar Integral. Van Nostrand.

[26] Potthoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve models. Biometrika, 51, p. 313.

[27] Rao, C.R. (1966). Covariance adjustment and related problems in multivariate analysis. Multivariate Analysis I (p. 87) (edited by P.R. Krishnuiah). Academic Press, New York.

[28] Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proc. 5th Berk. Symp. Math. Statist. and Prob. 1, p. 355-372.

[29] Rubin, D.B. (1976). Inference and missing data. Biometrika, 63, p. 581-592.

[30] Schwartz, R. (1967). Local minimax tests. Ann. Math. Statist., 38, p. 340-360.

[31] Stein, C. (1966). Notes on Multivariate Analysis. Stanford University. (Notes recorded by M.L. Eaton).

[32] Wijsman, R.A. (1967). Cross-sections of orbits and their application to densities of maximal invariants. Fifth Berk. Symp. Math. Statist. Prob. 1, p. 389-400, University of California.